

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования

«Воронежский государственный технический университет»

Кафедра управления

**ЭКОНОМИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ
(РАЗДЕЛ ЭКОНОМЕТРИКА)**

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к выполнению курсовых работ
для обучающихся всех направлений, специальностей,
форм обучения

Воронеж 2021

УДК 657 (075.8)
ББК 65.052.9(2)

Составитель ст. преп. Т. А. Свиридова

Экономико-статистические методы (раздел Эконометрика): методические указания к выполнению курсовых для обучающихся всех направлений, специальностей, форм обучения / ФГБОУ ВО «Воронежский государственный технический университет»; сост.: Т. А. Свиридова – Воронеж: Изд-во ВГТУ, 2021. - 38 с.

Основной целью является развитие навыков (полученных в рамках изучения курса): выбор и обоснование методов анализа представленной информации и формулирование выводов в рамках сделанного расчета.

Предназначены для обучающихся всех направлений, специальностей, форм обучения.

Методические указания подготовлены в электронном виде и содержатся в файле МУ_ЭСМ_КР_2021. pdf.

Табл. 1. Библиогр.: 10 назв.

УДК 657 (075.8)
ББК 65.052.9(2)

Рецензент – П. Н. Курочка д-р техн. наук, проф. кафедры управления ВГТУ

*Издается по решению редакционно-издательского совета
Воронежского государственного технического университета*

ОГЛАВЛЕНИЕ

1. Общие методические указания по изучению курса.....	4
2. Указания по выполнению курсовых работ	5
3. Теоретическая база в рамках изучения основных разделов дисциплины «Экономико-статистические методы».....	6
3.1. Предмет эконометрики. Парная регрессия и корреляция в эконометрических исследованиях.....	6
3.2. Множественная регрессия и корреляция.....	15
3.3. Связь между атрибутивными признаками знаками. Моделирование временных рядов.....	23
4. Задания для курсовой работы.....	31
Библиографический список.....	35
Приложение 1.....	36
Приложение 2.....	37

1. ОБЩИЕ МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ИЗУЧЕНИЮ КУРСА

Изучение предмета экономико-статистические методы (раздел эконометрика) имеет своей целью:

✓ овладение студентами статистической методологией и ее применение при всестороннем исследовании социально-экономических процессов, протекающих в организациях, на предприятиях, фирмах и в отраслях национальной экономики;

✓ овладение совокупностью математических методов, используемых для количественной оценки экономических явлений и процессов; обучение эконометрическому моделированию, т. е. построению экономико-математических моделей, параметры которых оцениваются средствами математической статистики;

✓ обучение эмпирическому выводу экономических законов; подготовку к прикладным исследованиям в области экономики.

Задачами дисциплины «Экономико-статистические методы» (раздел эконометрика) являются: получение студентами знаний и навыков формирования статистической информации, ее использования для получения обоснованной системы показателей, с помощью которых выявляются имеющиеся резервы роста эффективности производства и прогноз тенденций его развития;

✓ научить студентов использовать данные наблюдения для построения количественных зависимостей для экономических соотношений, для выявления связей, закономерностей и тенденций развития экономических явлений;

✓ выработать у студентов умение формировать экономические модели, основываясь на экономической теории или на эмпирических данных, оценивать неизвестные параметры в этих моделях, делать прогнозы и оценивать их точность, давать рекомендации по экономической политике и хозяйственной деятельности.

Дисциплина «Экономико-статистические методы» призвана сформировать широкий мировоззренческий горизонт будущего специалиста, а также закрепить и расширить методологическую и теоретическую базу, полученную по другим предметам.

Результатом освоения дисциплины является освоение соответствующих компетенций в рамках соответствующего учебного плана специальности и дисциплины:

Основными разделами изучаемой дисциплины «Экономико-статистические методы» являются:

1. Предмет эконометрики. Парная регрессия и корреляция в эконометрических исследованиях;
2. Множественная регрессия и корреляция;
3. Связь между атрибутивными признаками знаками. Моделирование временных рядов.

2. УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ КУРСОВЫХ РАБОТ

Курсовая работа по эконометрике имеет для подготовки студентов большое значение. Она содержит необходимый минимум задач, выполняя которые студент закрепляет полученные теоретические знания, осваивает методологию расчета статистических и эконометрических показателей. При обработке и анализе конкретных статистических данных студент знакомится с условиями применения математических методов и одновременно приобретает практические навыки по квалифицированному изложению результатов социально-экономических разработок.

Первоначальный вариант контрольной работы содержит 6 задач, которые формируются по следующей схеме. Каждый студент выполняет в своей курсовой работе 6 задач, полноценно поясняя и оформляя их. Разницей между каждым вариантом будет служить следующая последовательность действий. К каждому показателю, представленному в задаче (дополнительное пояснение есть к каждой задаче) прибавить число, сформированное двумя последними цифрами зачетной книжки, например, 20 (362+20 и т.д.) далее решить задачу с использованием вновь образованных данных). Важно отметить, что ваше число всегда должно быть двухзначным, если оно 02, 03 вы прибавляете не 2, 3, а берете предыдущее, например 506, следовательно, ваш вариант 56 и т.д. Работа, выполненная не по своим числовым данным, автоматически не зачитывается. Вариант может меняться по обоюдной договоренности преподаватель, студент в рамках более рационального подхода к учебному процессу (например, в случае, когда один и тот же вариант в группе встречается неоднократно).

Перед выполнением необходимо ознакомиться с изложенными ниже методическими указаниями.

Важно! Задания к курсовой работе могут корректироваться и определяться преподавателем индивидуально для каждого студента группы по обоюдной договоренности и согласию преподаватель-студент в рамках и целях оптимизации учебного процесса для более интенсивного и рационального освоения заявленных в рамках изучения курса компетенций, а также соизмеримо в часами и количеством ЗЕТ выделенных на дисциплину в рамках учебного плана и направления подготовки.

Для выполнения курсовой работы студентам можно воспользоваться не только рекомендуемой в настоящих методических указаниях литературой, но и литературой по выбору студента в рамках рассматриваемой дисциплины.

Курсовую работу следует выполнять на листах формата А4 с использованием текстового редактора MS Word и редактора электронных таблиц MS Excel.

Объем курсовой работы определяется студентом самостоятельно, исходя из логики изложения материала.

Для каждого из рассчитываемых в работе показателей следует привести полное наименование и формулу расчета. Если показатель рассчитан с использованием встроенных функций MS Excel, в качестве формулы расчета следует указать применяемую функцию MS Excel с перечислением всех

необходимых аргументов. Округлять рассчитанные значения показателей необходимо до трех или более десятичных знаков. Чем меньше округление, тем точнее результат.

Курсовая работа должна быть выполнена в сроки, установленные деканатом и преподавателем. Курсовая работа, содержащая ошибки, возвращается студенту на доработку. Курсовая работа, не содержащая ошибок, допускается преподавателем к защите.

Форма контроля: защита курсовой работы, включающая ответы на теоретические вопросы, связанные с тематикой курсовой работы, и пояснения по выполнению практических заданий.

3. ТЕОРЕТИЧЕСКАЯ БАЗА В РАМКАХ ИЗУЧЕНИЯ ОСНОВНЫХ РАЗДЕЛОВ ДИСЦИПЛИНЫ «ЭКОНОМИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ»

3.1. Предмет эконометрики. Парная регрессия и корреляция в эконометрических исследованиях

Предметом исследования эконометрики считаются массовые экономические процессы и явления. Эконометрика является наукой, которая эмпирически связана с выводом экономических законов, при этом используются данные или «наблюдения» для получения количественных зависимостей экономических соотношений. Экономические явления как предмет эконометрики (в отличие от экономической теории) рассматриваются в большей мере в количественном аспекте. Например, спрос на продукцию с ростом цен падает.

Методы теории корреляции позволяют определить количественную зависимость между различными техническими, технологическими, организационными и другими факторами, т.е. строить экономико-статистические модели.

Различают функциональную и корреляционную зависимости. Под функциональной понимается такая зависимость, когда с изменением одного фактора изменяется другой, одному значению независимого фактора обычно соответствует только одно значение зависимого фактора. Корреляционная зависимость - это такая зависимость, при которой изменение одной случайной величины вызывает изменение среднего значения другой. Конкретных же значений зависимого переменного, соответствующих одному значению независимого, может быть несколько. Корреляционные зависимости могут быть установлены только при обработке большого количества наблюдений. При корреляционном анализе решаются следующие задачи: устанавливается наличие корреляции (связи) между величинами; устанавливается формула линии связи (линии регрессии); определяются параметры линии регрессии; определяются значимость установленной зависимости и достоверность отдельных параметров.

Наличие корреляции приближенно может быть определено путем визуального анализа поля корреляции. Корреляционным полем называют нанесенные на график в определенном масштабе точки, соответствующие одновременным значениям двух величин.

Тесноту связи между двумя величинами можно определить визуально по соотношению короткой и продольной осей эллипса рассеяния наблюдений, нанесенных на поле корреляции. Чем больше отношение продольной стороны к короткой, тем связь теснее.

Более точно теснота связи характеризуется коэффициентом корреляции r . Коэффициент корреляции лежит в пределах $0 \leq |r| \leq 1$. В случае, если $r=0$, то линейной связи нет. Если $|r|=1$, то между двумя величинами существует функциональная связь. При положительном r наблюдается прямая связь, т.е. с увеличением независимого переменного увеличивается зависимое. При отрицательном коэффициенте наблюдается обратная связь - с увеличением независимого переменного зависимое переменное уменьшается.

Коэффициент корреляции определяется по формуле

$$r = \frac{N \sum_{i=1}^N x_i \cdot y_i - \sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i}{\sqrt{N \cdot \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \cdot \sqrt{N \cdot \sum_{i=1}^N y_i^2 - (\sum_{i=1}^N y_i)^2}},$$

где x и y - текущие значения наблюдаемых величин; N - число наблюдений.

Для численного выражения параметров линии регрессии, выражающих связь между двумя величинами, обычно применяется метод наименьших квадратов. Сущность этого метода состоит в том, что выбирается такая линия, при которой сумма квадратов разностей между фактическими наблюдениями зависимой переменной и расчетными значениями, полученными по регрессионной формуле, минимальна

$$S = \sum (y - \tilde{y})^2 \rightarrow \min,$$

где \tilde{y} - расчетное значение зависимого переменного по регрессивной формуле.

Допустим, $\tilde{y} = a + bx$. Тогда

$$S = \sum (y - a - bx)^2 \rightarrow \min$$

Возьмем частные производные по a и по b и приравняем их к нулю:

$$\frac{\partial S}{\partial a} = -2 \sum (y - a - bx) = 0$$

$$\frac{\partial S}{\partial b} = -2 \sum (y - a - bx) = 0$$

Полученную систему обычно преобразуют

$$\begin{aligned} Na + b \sum x &= \sum y \\ a \sum x + b \sum x^2 &= \sum xy \end{aligned}$$

Решив систему относительно b и a , получим формулы

$$b = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2},$$

$$a = \bar{y} - b\bar{x}.$$

При линейной корреляции коэффициент корреляции r является не только критерием тесноты связи, но и критерием точности аппроксимации (подбора формулы, выражающей зависимость).

Далеко не всегда данные статистических наблюдений можно аппроксимировать в виде линейной зависимости. Очень часто оказывается, что линейная зависимость не дает необходимой тесноты связи. В этом случае аппроксимацию проводят с помощью зависимостей, отличных от линейной. С этой целью чаще всего используются

- * степенная;
- * логарифмическая;
- * параболическая;
- * многочлен степени n ;
- * зависимости периодического вида.

Для определения параметров нелинейной зависимости пользуются процедурой метода наименьших квадратов, но предварительно проводят линеаризацию. С этой целью, как правило, производят логарифмирование, например, зависимость степенного вида представляется аналитической формулой

$$\tilde{y} = ax^b,$$

которая после логарифмирования будет представлена в виде

$$\lg \tilde{y} = \lg a + b \lg x,$$

параметры lga и b находятся по методу наименьших квадратов по формуле

$$lga = \frac{\sum lgy \sum (lgx)^2 - \sum lgx \sum lgy lgx}{N \sum (lgx)^2 - (\sum lgx)^2},$$

$$b = \frac{N \sum lgx lgy - \sum lgx \sum lgy}{N \sum (lgx)^2 - (\sum lgx)^2}.$$

Аналогично можно получить выражение для параметров остальных типов кривых.

Логарифмическая зависимость выражается формулой вида:

$$y = a + b \lg x.$$

Расчетные формулы для определения параметров a и b имеют вид:

$$a = \frac{A}{D}; b = \frac{B}{D},$$

где

$$A = \sum y \sum (lgx)^2 - \sum lgx \sum y lgx;$$

$$B = N \sum y lgx - \sum lgx \sum y;$$

$$D = N \sum (lgx)^2 - (\sum lgx)^2$$

В виде параболы второго порядка зависимость выражается формулой

$$y = a + bx + cx^2.$$

Если степень независимого переменного равна трем, то эта парабола третьего порядка и т.д. Линейная зависимость также является частным случаем многочлена.

Аппроксимация (определение параметров) параболической кривой осуществляется методом наименьших квадратов. В целевую функцию метода наименьших квадратов $\sum (y - \tilde{y})^2 \rightarrow \min$ вместо расчетных значений y подставим правую часть уравнения:

$$S = \sum (y - bx - cx^2) \rightarrow \min.$$

Возьмем частные производные от этого выражения по a, b и c:

$$\frac{\partial S}{\partial a} = -2 \sum (y - a - bx - cx^2) = 0;$$

$$\frac{\partial S}{\partial b} = -2 \sum (y - a - bx - cx^2)x = 0;$$

$$\frac{\partial S}{\partial c} = -2 \sum (y - a - bx - cx^2)x^2 = 0.$$

Получим систему нормальных или ортогональных уравнений, которая после несложных преобразований примет вид:

$$\begin{aligned} Na + b \sum x + c \sum x^2 &= \sum y; \\ a \sum x + b \sum x^2 + c \sum x^3 &= \sum yx; \\ a \sum x^2 + b \sum x^3 + c \sum x^4 + \sum yx^2 &. \end{aligned}$$

Решая систему любым известным методом, находим параметры параболы a, b и c.

Параметры параболы можно определить из выражений:

$$\left. \begin{aligned} a &= \frac{D_a}{D}; \\ b &= \frac{D_b}{D}; \\ c &= \frac{D_c}{D}, \end{aligned} \right\}$$

где D - главный определитель системы линейных уравнений; D_a - определитель системы уравнений, в котором столбец коэффициентов при a заменяют столбцом свободных членов; D_b - определитель системы, в котором столбец коэффициентов при b заменен столбцом свободных членов; D_c - определитель системы, в котором столбец коэффициентов при «с» заменен столбцом свободных членов.

Определители матрицы можно расписать в виде следующих выражений:

$$\begin{aligned}
D &= N \sum x^2 \sum x^4 + \sum x \sum x^3 \sum x^2 + \sum x \sum x^2 \sum x^3 - \\
&\quad - \sum x^2 \sum x^2 \sum x^2 - (\sum x)^2 \sum x^4 - N(\sum x^3)^2, \\
D_a &= \sum y \sum x^2 \sum x^4 + \sum x^3 \sum x \sum x^2 y + \sum xy \sum x^2 \sum x^3 - \\
&\quad - \sum x^2 y (\sum x^2)^2 - \sum xy \sum x \sum x^4 - \sum y (\sum x^3)^2, \\
D_b &= N \sum xy \sum x^4 + \sum y \sum x^3 \sum x^2 + \sum x \sum x^2 \sum x^2 y - \\
&\quad - N \sum x^3 \sum x^2 y - \sum x^4 \sum x \sum y - (\sum x^2)^2 \sum xy, \\
D_c &= N \sum x^2 \sum x^2 y + \sum x \sum x^3 \sum y + \sum x \sum x^2 \sum xy - \\
&\quad - N \sum x^3 \sum xy - \sum x^2 y \sum x^2 \sum y - (\sum x^2)^2 \sum y.
\end{aligned}$$

Кривые периодического вида могут найти широкое распространение при аппроксимации зависимости многих экономических явлений во времени. Например, такими кривыми выражаются влияния сезонных факторов на организацию строительства и материально-технического обеспечения. Наблюдения времени при этом можно представить в виде равноотстоящих, например, x , выраженных в радианах или градусах. Если взять период времени, равный году, и провести ежемесячные наблюдения какого-либо экономического показателя, то время как аргумент может быть записан в виде

$$\frac{1}{12} 2\pi; \frac{2}{12} 2\pi; \frac{3}{12} 2\pi; \dots$$

В течение года можно получить 12 наблюдений экономического показателя y_1, y_2, \dots, y_{12} . Тогда зависимость величины «у» от времени можно выразить уравнением

$$\tilde{y} = a_0 + \sum_{k=1}^m (a_k \cos kx + b_k \sin kx),$$

где $k=1, 2, 3, \dots, m$ - заданное число гармоник многочлена;

a_0, a_k , и b_k - коэффициенты линии регрессии, общим числом $2m+1$.

Если $N > 2m+1$, то коэффициенты a_k и b_k находятся по методу наименьших квадратов, который дает следующие выражения для коэффициентов линии регрессии:

$$\left. \begin{aligned} a_0 &= \frac{1}{N} \sum_{i=1}^N y_i, \\ a_k &= \frac{2}{N} \sum_{i=1}^N y_i \cos kx_i, \\ b_k &= \frac{2}{N} \sum_{i=1}^N y_i \sin kx_i. \end{aligned} \right\}$$

Оценка точности аппроксимации криволинейной зависимостью производится при помощи корреляционного отношения:

$$\eta = \sqrt{1 - \frac{\sum (y - \tilde{y})^2}{\sum (y - \bar{y})^2}}$$

Корреляционное значение всегда $0 \leq \eta \leq 1$, оно всегда положительно. Если $\eta > r$, то кривая точнее аппроксимирует зависимость, чем прямая; для прямой $r = \eta$.

Дополнительной оценкой точности аппроксимации, часто применяемой при оценке нелинейной корреляции, является средняя относительная ошибка аппроксимации, которая определяется по формуле

$$\bar{\varepsilon} = \frac{1}{N} \sum \left| \frac{y - \tilde{y}}{y} \right| \cdot 100.$$

Коэффициент корреляции r , рассчитанный по выборочным данным, может не совпасть с истинным коэффициентом корреляции, соответствующим генеральной совокупности ρ .

Среднеквадратичное отклонение приближенно определяется по формуле

$$\sigma_r \cong \frac{1 - r^2}{\sqrt{N - 1}}.$$

При больших выборках можно предположить, что коэффициент корреляции распределен по нормальному закону, тогда можно утверждать, что

$$P\{r - x_p \sigma_r \leq \rho \leq r + x_p \sigma_r\} = \Phi(x)$$

Особенно интересна проверка так называемой нулевой гипотезы. Известно, что если коэффициент корреляции по модулю больше 0, то между двумя случайными величинами имеется связь. Однако r , определенный по частичной выборке, отличается от истинного коэффициента корреляции ρ .

Может быть, что при $|r| > 0$, $\rho = 0$, тогда связь, установленная по частичной выборке, в генеральной совокупности, отсутствует. Для ответа на вопрос, есть ли связь в генеральной совокупности, осуществляют проверку на значимость (коэффициент корреляции существенно отличен от нуля) коэффициента корреляции. Если

$$|r| > \frac{x_p}{\sqrt{N-1}}$$

то с заданной вероятностью P можно утверждать, что коэффициент существенно отличен от нуля (это означает, что нулевая гипотеза отвергается) и рассматриваемая связь в генеральной совокупности существует. Однако при малых объемах статистического материала гипотеза о нормальном распределении коэффициента корреляции, как правило, не подтверждается. При небольшом числе испытаний вопрос о значимости коэффициента корреляции рассматривается с использованием t -критерия Стьюдента. При этом определяется расчетное значение критерия по формуле

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{N-2}$$

, где $N-2$ - число степеней свободы.

Теоретическое значения определяется по таблице распределения Стьюдента (прил. II). Если $t_{\text{рас.}} > t_{\text{табл.}}$ при заданном уровне значимости, то предположение о нулевом значении коэффициента корреляции в генеральной совокупности не подтверждается.

При аппроксимации корреляционной зависимости полученная линия регрессии отвечает только частичной выборке, то есть тем данным, которые были использованы при статистической обработке. Для распространения этой зависимости на генеральную совокупность необходимо оценить значение коэффициентов регрессии, так как может оказаться, что при условии неравенства коэффициента регрессии нулю истинный коэффициент регрессии, отражающий генеральную совокупность, будет нулевым. В этом случае прогнозировать по полученной кривой нельзя. Значимость отдельных коэффициентов определяется при помощи t -критерия Стьюдента.

Для оценки значимости коэффициентов регрессии расчетное значение t -критерия Стьюдента определяется по формулам

$$t_{a_i} = \frac{a_i \sqrt{(N-n) \cdot \sum_{i=1}^N (x - \bar{x})^2}}{\sqrt{\sum_{i=1}^N (y - \bar{y})^2}}$$

$$t_{a_i} = \frac{a_i}{D_{\text{ост}} \cdot \sqrt{c_{ii}}},$$

где a_i - коэффициент регрессии при i - м члене уравнения регрессии;
 n - число коэффициентов регрессии;

$D_{\text{ост}} = \frac{\sum (y - \tilde{y})^2}{N - n - 1}$ - остаточная дисперсия;

c_{ii} - диагональный элемент обратной матрицы.

По первой формуле определяется значение t - критерия при одной переменной, а по второй формуле - при множественной корреляции.

При этом если $t_{a_i} \geq t_{\text{табл}}$, то коэффициент корреляции существенно отличен от нуля, а следовательно, этот коэффициент имеет значение отличное от нуля и в генеральной совокупности.

Оценка коэффициентов регрессии при помощи t - критерия Стьюдента применяется только для линейных корреляционных связей. Но так как при помощи метода наименьших квадратов путем линеаризации определяются линейные коэффициенты регрессии, то t - критерий может применяться также для любого вида функции в линеаризованной форме.

Значимость уравнения регрессии определяется возможностью надежно прогнозировать средние значения исследуемой величины. Уравнение прогноза получено на основе частичной совокупности, но истинная зависимость, свойственная генеральной совокупности, может существенно отличаться от полученного соотношения. Для изучения степени соответствия полученного уравнения регрессии истинному соотношению следующему из генеральной совокупности используют F - критерий Фишера, определяемый по формуле

$$F = \frac{D_Y}{D_{\text{ост}}}$$

где $D_Y = \frac{\sum (y - \bar{y})^2}{N - 1}$ - дисперсия фактических значений зависимого переменного; $N - n - 1 = f_1$; $N - 1 = f_2$ - число степеней свободы.

По числу степеней свободы, задавшись вероятностью, можно определить табличное значение критерия Фишера, значения для которых приведены в прил. III. Если $F \geq F_{\text{табл}}$, то уравнение регрессии считается значимым, т.е. уравнение будет давать достаточно надежные прогнозы и может быть использовано.

Уравнение регрессии из-за вероятностного характера имеет некоторую случайную компоненту h , на величину которой могут отличаться значения зависимой переменной от ее истинных значений

Величина « h » отражает влияние неучтенных факторов и несоответствие частичной совокупности, по которой определялось уравнения регрессии генеральной совокупности. Для надежного прогнозирования необходимо определить доверительный интервал исследуемой величины y .

Если предположить, что величина h является случайной величиной, распределенной по нормальному закону, то истинное значение случайной величины « y » будет находиться в интервале

$$\tilde{y} - x_a \cdot \sigma_h \leq y_{\text{ист}} \leq \tilde{y} + x_a \cdot \sigma_h$$

где $\sigma_h = \sqrt{D_{\text{ост}}^2}$;

x_a - аргумент, характеризующий вероятность попадания случайной величины в заданные пределы.

3.2. Множественная регрессия и корреляция

Большинство закономерностей, изучаемых управлением, имеет достаточно сложный характер и зависит от нескольких факторов, степень влияния каждого на изучаемую величину неизвестна. Первым этапом в исследовании таких зависимостей является создание и анализ парных корреляционных моделей, характеризующих парные зависимости между изучаемой величиной и факторным признаком. Такие модели не отражают всю причинно-следственную связь в изучаемом явлении, но могут дать какие-то первичные сведения о поведении и характере изменения изучаемой величины от каждого из факторов. Но для анализа сложных явлений необходимо изучить его в максимально полной взаимосвязи всех существенных факторных признаков и оценить степень влияния каждого из них на исследуемое явление. Для исследования подобных зависимостей используются модели множественной корреляции.

Для построения множественной корреляционной модели необходимо определить, какая величина будет являться зависимой, т.е. результирующим показателем, а какие величины будут являться независимыми, иначе называемые факторными признаками (факторами). Задача моделирования состоит в выявлении количественной связи между факторами и результирующим показателем.

Фактор, включаемый в модель, должен соответствовать следующим требованиям: иметь количественное выражение; между фактором и результирующим показателем должна быть логическая, причинная связь; между фактором и результирующим показателем должна быть статистическая

связь; факторы не должны быть тесно связаны между собой, т.е. между факторами не должно быть мультиколлинеарности.

Последнее обстоятельство означает, что парный коэффициент корреляции между двумя факторами не должен быть больше 0,85. Имеются и другие подходы к преодолению явления мультиколлинеарности: можно рекомендовать включать в многофакторные модели те факторы, коэффициент корреляции между которыми не оказался значимым при вероятности 0,9, т.е. для них подтвердилась нулевая гипотеза.

Так же как при парной корреляции, простейшей формой выражения множественной зависимости является линейная зависимость вида:

$$\tilde{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n,$$

где \tilde{y} - результирующий признак; x_i - факторы; a_0 - свободный член уравнения регрессии; a_i - коэффициенты при факторах. Линейная зависимость является частным случаем параболической зависимости, имеющей выражение:

$$\tilde{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{12}x_1x_2 + a_{13}x_1x_3 + \dots + a_{1n}x_1x_n + \dots, \quad (4.2)$$

где x_i^2 - квадраты значений факторов; x_ix_j - попарные произведения всевозможных комбинаций факторов; a_{ig} - коэффициент регрессии при попарном произведении факторов (i, j - соответственно номера факторов).

$$\text{Степенная зависимость } \tilde{y} = Ax_1^{b_1}x_2^{b_2} \dots x_n^{b_n} = A \prod_{i=1}^{i=n} x_i^{b_i},$$

где b_i - показатель степени при каждом i -том факторе; заметим, что коэффициенты b_i могут принимать целые и дробные значения с положительным знаком, в последнем случае кривая принимает вид гиперболы; A - числовой коэффициент.

$$\text{Показательная зависимость } \tilde{y} = a_1^{b_1x_1} a_2^{b_2x_2} \dots a_n^{b_nx_n} = \prod_{i=1}^{i=n} a_i^{b_ix_i},$$

где a_i и b_i - коэффициенты регрессии, принимающие любые вещественные значения.

Частным случаем показательной зависимости является экспоненциальная зависимость

$$\tilde{y} = e^{f(x_1, \dots, x_n)} = \exp[f(x_1, \dots, x_n)], \quad (4.5)$$

где $f(x_i)$ - любые функции факторов x_i при $i=1, 2, \dots, n$.

всего решать матричным способом. В этом случае система уравнений запишется в виде матричного уравнения

$$(X^*X)A = X^*Y,$$

где X - матрица исходных статистических данных по независимым переменным - факторам;

X^* - транспонированная матрица исходных данных;

Y - матрица - столбец значений зависимой переменной - отклика;

A - матрица - столбец искомых коэффициентов уравнения регрессии.

Эти матрицы представляются в следующем виде:

$$A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ \dots \\ \dots \\ a_n \end{pmatrix} \quad X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & X_{Nj} & \dots & X_{Nn} \end{pmatrix} \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ \dots \\ \dots \\ Y_N \end{pmatrix}$$

Транспонированная матрица получается из исходной матрицы X путем замены строк на столбцы. В этом случае общий вид транспонированной матрицы имеет следующий вид:

$$X^* = \begin{pmatrix} X_{11} & X_{21} & \dots & X_{i1} & \dots & X_{N1} \\ X_{12} & X_{22} & \dots & X_{i2} & \dots & X_{N2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{1j} & X_{2j} & \dots & X_{ij} & \dots & X_{Nj} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{1n} & X_{2n} & \dots & X_{in} & \dots & X_{Nn} \end{pmatrix}$$

Здесь n - число факторных признаков в корреляционной модели;

N - объем статистического материала.

Решение матричного уравнения находится с помощью обращения матриц и записывается в виде

$$A = (X^*X)^{-1}(X^*Y)$$

где $(\mathbf{X}^* \mathbf{X})^{-1}$ - матрица, обратная к $(\mathbf{X}^* \mathbf{X})$.

Найти матрицу \mathbf{A}^{-1} , обратную матрице \mathbf{A} составленную из элементов

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

можно следующим образом:

- * вычислить определитель этой матрицы D ;
- * определить алгебраические дополнения A_{ij} каждого элемента определителя матрицы a_{ij} . Напомним, что алгебраические дополнения определяются вычеркиванием i -ой строки и j -ого столбца;
- * составить из алгебраических дополнений матрицу $\tilde{\mathbf{A}}$;
- * составить из $\tilde{\mathbf{A}}$ транспонированную матрицу $\tilde{\mathbf{A}}^*$ (для чего необходимо поменять строки и столбцы местами);
- * составить обратную матрицу $\mathbf{A}^{-1} = \tilde{\mathbf{A}}^* / D$.

Если размерность матрицы $(\mathbf{X}^* \mathbf{X})$ достаточно большая (как правило больше 4), то ее обращение связано с определенными трудностями, объясняющимися большим объемом вычислительной работы, и эта работа прodelывается с помощью ЭВМ; в этом случае возможно возникновение проблем, связанных с вычислительной устойчивостью применяемого алгоритма обращения, что порождается плохой обусловленностью исходной матрицы, то есть определитель близок к нулю.

Реальная связь изучаемых факторов, как правило, имеет зависимость, существенно отличную от линейной (зачастую такая зависимость может быть представлена в виде линии только на очень ограниченном интервале изменения факторных признаков), а поэтому для построения адекватных корреляционных моделей требуется использование нелинейных аппроксимаций.

Наиболее универсальной моделью, аппроксимирующей нелинейную связь, является многочлен n -ой степени. Из теории интерполяции известно, что полиномом произвольной степени можно описать практически любую зависимость с любой степенью точности, но такое представление связано с огромным объемом вычислительной работы и вряд ли применимо на практике. Обычно при определении максимально возможной степени полинома ориентируются на объем статистического материала, имеющегося в распоряжении исследователя. При этом следует иметь в виду, что число членов полинома должно быть меньше числа наблюдений в противном случае происходит переход к функциональной зависимости. Для ориентировочных оценок можно использовать соотношение

$$\frac{N}{n+1} > 8.$$

При аппроксимации полиномом n -ой степени уравнение регрессии должно быть предварительно линеаризовано. С этой целью все нелинейные члены, входящие в уравнение регрессии, заменяются новыми, дополнительными факторными признаками. Это приводит к увеличению размерности задачи, но делает доступным аппарат исследования линейных моделей.

Алгоритм построения многофакторной корреляционной модели заключается в следующем:

- * задать максимально возможно высокую степень полинома аппроксимации, записать уравнение регрессии в общем виде;
- * осуществить линеаризацию исходного уравнения регрессии и записать систему разрешающих линейных алгебраических уравнений в матричной форме;
- * найти коэффициенты уравнения регрессии;
- * проверить значимость уравнения регрессии по критерию Фишера; если уравнение регрессии оказалось незначимым, то требуется увеличить степень многочлена и повторить все вычисления сначала;
- * проверить значимость коэффициентов уравнения регрессии с помощью t -критерия Стьюдента, отбрасывая те из них, которые незначимы, и вновь повторяя построение модели уже с меньшим количеством факторов.

Построение многофакторных моделей с использованием нелинейных форм аппроксимации, как правило, связано с большим объемом вычислений и осуществляется с помощью ЭВМ. В этой связи очень важен вопрос о снижении объемов вычислительной работы, что может быть достигнуто за счет применения нетрадиционных форм аппроксимации. Одним из таких подходов является метод Брандона, когда уравнение регрессии задается в виде произведения, каждый сомножитель которого представлен функцией только одного факторного признака. Таким образом, уравнение принимает вид

$$\tilde{y} = c \prod_{i=1}^n f_i(x_i)$$

Здесь $y = f_i(x_i)$ может иметь любой вид;

c - постоянная величина, равная среднему значению \bar{y} .

Алгоритм работы по методу Брандона заключается в следующем:

- * вычисляется среднее значение y и преобразуются значения для каждого наблюдения по формуле (j - порядковый номер наблюдения)

$$y_{oj} = \frac{y_j}{\bar{y}} ;$$

- * выбирается вид зависимости y_0 от x_1 и по методу наименьших квадратов определяются параметры формулы $\tilde{y}_0 = f_1(x_1)$;
- * вычисляются значения функции $f_1(x_1)$ и определяется остаточный показатель y_1 для каждого наблюдения по формуле

$$y_1 = \frac{y_0}{f_1(x_1)}$$

в предположении, что y_1 зависит от x_2, x_3, \dots, x_n и не зависит от x_1 ;

- * определяется корреляционная формула зависимости y_1 от x_2 ;
- * находится условный показатель, не зависящий от x_1 и x_2

$$y_2 = \frac{y_1}{f_2(x_2)}$$

Такие преобразования результирующего показателя осуществляются до тех пор, пока не будет определена вся последовательность функций, входящих в произведение.

Данный алгоритм дает существенное снижение объемов вычислений и их качественное упрощение: уже не требуется решать алгебраических систем большой размерности.

Количественно тесноту связи при множественной корреляции можно оценить с помощью множественного (совокупного) коэффициента корреляции R . Для расчета совокупного коэффициента корреляции необходимо определить парные коэффициенты корреляции r_{0i} между всеми факторами x_i , входящими в модель, и результирующим показателем y и все парные коэффициенты корреляции между факторами. Все коэффициенты корреляции записываются в квадратную симметричную матрицу

$$\begin{bmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_n} \\ r_{yx_2} & r_{x_1x_2} & 1 & r_{x_2x_3} & \dots & r_{x_2x_n} \\ r_{yx_3} & r_{x_1x_3} & r_{x_2x_3} & 1 & \dots & r_{x_3x_n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_1x_n} & r_{x_2x_n} & r_{x_3x_n} & \dots & 1 \end{bmatrix}$$

Множественный коэффициент корреляции определяется по формуле

$$R = \sqrt{1 - \frac{D}{D_{11}}}$$

где D - определитель матрицы парных коэффициентов корреляции;

D_{11} - определитель той же матрицы с вычеркнутыми первой строкой и первым столбцом, т.е. определитель матрицы парных коэффициентов корреляции между факторами.

Для случая зависимости от двух факторов

$$R_{y/x_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 + 2r_{yx_1}r_{yx_2}s_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

Для определения влияния только одного i -того фактора на результирующий показатель с исключением влияния других факторов используется частный коэффициент корреляции

$$r_{y/x_1x_2\dots x_n} = \frac{D_{1i}}{\sqrt{D_{11}D_{ii}}}$$

где D_{1i} - определитель матрицы с вычеркнутой первой строкой и i -тым столбцом; D_{ii} - определитель матрицы с вычеркнутой i -той строкой и i -тым столбцом.

При множественной корреляции от двух факторов коэффициент частной корреляции первого фактора

$$r_{y/x_1x_2} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}}$$

а коэффициент частной корреляции для второго фактора

$$r_{y/x_2x_1} = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}}$$

Частный коэффициент корреляции отражает « чистое » влияние фактора на результирующий показатель и отличается от коэффициента парной корреляции r_{yx} .

При линейной форме связи множественный коэффициент корреляции является оценкой точности аппроксимации и равен корреляционному отношению η ; при нелинейных формах связи для оценки точности

аппроксимации (оценки адекватности модели) применяются корреляционное отношение η и ошибка аппроксимации ε . Эти оценки определяются так же, как и при парной корреляции.

Для оценки влияния отдельных факторов на результативный признак используются коэффициенты эластичности, показывающие уровень изменения результативного показателя в случае изменения факторного признака на 1 % при неизменных значениях других факторов. Коэффициенты эластичности определяются по формуле.

$$\varepsilon_i = \frac{x_i}{y} \cdot \frac{\partial \tilde{y}}{\partial x_i},$$

где \tilde{y} - теоретическое уравнение регрессии.

3.3. Связь между атрибутивными признаками знаками. Моделирование временных рядов

Использование регрессионного и корреляционного анализа требует, чтобы все признаки были количественно измеренными. Методы КРА, основанные на использовании количественных параметров распределения (средние величины, дисперсия), называют параметрическими методами.

Вместе с тем, особенно при проведении социологических исследований, возникает потребность оценки тесноты связи между качественными (атрибутивными) признаками. Проблему оценки тесноты связи между атрибутивными признаками решают непараметрические методы. Сфера их использования значительно шире в сравнении с параметрическими методами, потому что не требуется использования условия нормального распределения результативной переменной, не ставится задача представления зависимости между атрибутивными признаками соответствующим уравнением. Здесь речь идет только о наличии установлении связи и измерения его тесноты.

Взаимосвязь между атрибутивными признаками анализируются посредством таблиц взаимной сопряженности. Они описывают комбинационные распределения совокупности по факторному признаку x и результативному y . Например, результаты социологического опроса населения относительно намерений принять участие на рынке ценных бумаг: распределение респондентов опроса по возрасту рассматривается как факторный признак x , а их распределение по склонности к риску как результативный признак y . Таблицы взаимной сопряженности могут иметь различную размерность. Простейшая размерность – 2×2 (таблица «четырёх полей»), когда по альтернативному признаку («да» – «нет», «хорошо» – «плохо» и т.д.) выделяются 2 группы.

При наличии стохастической связи оценка его тесноты базируется на отклонениях фактических частот f_{ij} от F_{ij} , пропорциональных итоговым частотам:

$$F_{ij} = \frac{f_{i0} f_{0j}}{n}$$

, где f_{i0} - суммарные частоты по признаку x ; f_{0j} - суммарные частоты по признаку y ; n - объем совокупности. Очевидно, что

$$m = \sum_{i=1}^{m_x} f_{i0} = \sum_{j=1}^{m_y} f_{j0}$$

, где m_x, m_y – соответственно количество групп по признакам x и y .

Абсолютную величину отклонений фактических частот f_{ij} от пропорциональных F_{ij} ($f_{ij} - F_{ij}$) характеризуют статистическим критерием χ^2 (хи-квадрат).

$$\chi^2 = \sum_{i=1}^{m_x} \sum_{j=1}^{m_y} \frac{(f_{ij} - F_{ij})^2}{F_{ij}} = n \left(\sum_{i=1}^{m_x} \sum_{j=1}^{m_y} \frac{f_{ij}^2}{f_{i0} f_{0j}} - 1 \right)$$

Из-за отсутствия стохастической связи $\chi^2 = 0$. Для вывода о тесноте связи теоретическое значение χ^2 сравнивается с табличным $\chi^2_{\text{табл}}$. Последний выбирается из справочных математических таблиц критерия «хи»-квадрат в зависимости от принятого уровня значимости α (0,01 или 0,05), и степеней свободы $k = (m_x - 1)(m_y - 1)$. При $\chi^2 > \chi^2_{\text{табл}}$ делают вывод о наличии тесной связи между признаками x и y .

Однако выводы о зависимости, сделанные «на глаз», часто могут быть ненадежными (ошибочными), поэтому они должны подкрепляться определенными статистическими критериями, например критерием Пирсона χ^2 . Он позволяет судить о случайности (или неслучайности) распределения в таблицах взаимной сопряженности, а, следовательно, и об отсутствии или наличии зависимости между признаками группировки в таблице. Чтобы воспользоваться критерием Пирсона χ^2 , в таблице взаимной сопряженности наряду с эмпирическими часто там записывают теоретические частоты, рассчитываемые исходя из предположения, что распределение внутри таблицы случайно и, следовательно, зависимость между признаками группировки отсутствует. То есть считается, что распределение частот в каждой строке (столбце) таблицы пропорционально распределению частот в итоговой строке (столбце). Поэтому теоретические частоты по строкам (столбцам) рассчитывают пропорционально распределению единиц в итоговой строке (столбце).

Относительной мерой тесноты стохастической связи между признаками служат также:

коэффициент взаимной сопряженности Чупрова

$$C = \sqrt{\frac{\chi^2}{n\sqrt{(m_x - 1)(m_y - 1)}}$$

Но, следует отметить, что рассчитывать коэффициент Чупрова для таблицы «четырёх полей» не рекомендуется, так как при числе степеней свободы $\nu=(2-1)(2-1)=1$ он будет больше коэффициента Пирсона (в нашем примере $C=0,54$). Для таблиц же большей размерности всегда $C < \chi^2$.

Коэффициент взаимной сопряженности Крамера (при $T_x \neq T_y$) характеризует меру связи двух номинальных переменных на основе критерия хи-квадрат. Применяется к таблицам сопряженности произвольной размерности

$$C = \sqrt{\frac{\chi^2}{n\sqrt{(m_{min} - 1)}}$$

где m_{min} — минимальное число групп (m_x или m_y).

Значение коэффициента колеблется от 0 до 1 и теснота связи тем сильнее, чем более близко C к 1.

Достаточно часто в практике статистических исследований анализируются связи между альтернативными признаками, которые представлены группами с противоположными (взаимоисключающими) характеристиками. Тесноту связи в этом случае можно оценивать посредством коэффициента ассоциации Д. Юла и коэффициента контингенции К. Пирсона.

Для расчета указанных коэффициентов измерения тесноты связи между альтернативными признаками используется таблица взаимной сопряженности в виде корреляционной таблицы, которая носит название «четырёхклеточной табл.».

Таблица

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$a+b+c+d$

При использовании табл. с частотами a, b, c, d коэффициент ассоциации (K_a) вычисляется по формуле:

$$K_a = \frac{ad - bc}{ad + bc}$$

При $K_a > 0,3$ между изучаемыми качественными признаками существует корреляционная связь.

В случаях, когда один из показателей четырехклеточной таблицы отсутствует, величина коэффициента ассоциации будет равняться единице, что

дает завышенную оценку тесноты связи между признаками. В этом случае необходимо рассчитывать коэффициент контингенции (K_k):

$$K_k = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$$

Коэффициент контингенции находится в диапазоне от -1 к +1. Чем более близко K_k к (+1) или (-1), тем теснее связь между изучаемыми признаками. Коэффициент контингенции всегда меньше коэффициента ассоциации.

Для определения связи, как между количественными, так и качественными признаками при условии, что значения этих признаков упорядочены по степени уменьшения или увеличения (ранжированные), может быть использован коэффициент корреляции рангов Спирмена. Рангами называют числа натурального ряда, которые представляются в баллах по определенным критериям элементов совокупности. При этом ранжирование проводится по каждому признаку отдельно: первый ранг предоставляется наименьшему значению признака, последний - наибольшему. Количество рангов равняется объему совокупности. Преимуществом этого подхода является то, что при отсутствии требования нормального распределения ранговые оценки тесноты связи целесообразно использовать для совокупности небольшого объема.

Показатель ранговой корреляции - коэффициент корреляции рангов Спирмена — рассчитывается по формуле:

$$\rho = 1 - \frac{6 \sum_{j=1}^n d_j^2}{n(n^2 - 1)}$$

, где d_j — разность между рангами по одному и другому признаку ($d_j = R_{xj} - R_{yj}$); n - количество единиц в ряду. Если $d_j = 0$ $\rho = 1$ — существует тесная прямая связь. Если первому рангу по размеру одного признака соответствует последний ранг по размеру второго признака, второму рангу - предпоследний ранг второго признака и т. п., то $\rho = -1$ и существует тесная обратная связь. Если значения ρ близко к нулю, то связь слабая или ее вообще нет. Для предварительной оценки тесноты связи между атрибутивными признаками используются также такие характеристики, как коэффициенты Фехнера и Кендэла.

Коэффициент Фехнера относится к простейшим показателям степени тесноты связи и иногда называют коэффициентом корреляции знаков, который был предложен немецким ученым Г. Фехнером (1801 — 1887). Этот показатель основан на оценке степени согласованности направлений отклонений индивидуальных значений факторного и результативного признаков от соответствующих средних. Для его расчета вычисляют средние значения результативного и

факторного признаков, а затем проставляют знаки отклонений для всех значений взаимосвязанных пар признаков.

Обозначим через n_a число совпадений знаков отклонений индивидуальных величин от средней, через n_b — число несовпадений таковых отклонений. Формула коэффициента Фехнера записывается так:

$$K_f = (n_a - n_b) / (n_a + n_b).$$

Коэффициент Фехнера может принимать различные значения в пределах от -1 до $+1$. Если знаки всех отклонений совпадут, то $n_b = 0$ и тогда коэффициент будет равен 1 , что свидетельствует о возможном наличии прямой связи. Если же знаки всех отклонений будут разными, тогда $n_a = 0$ и коэффициент Фехнера будет равен -1 , это дает основание предположить наличие обратной связи.

Как видно из приведенной формулы для расчета коэффициента Фехнера, величина этого показателя не зависит от величины отклонений факторного и результативного признаков от соответствующей средней величины. Поэтому нельзя говорить о степени тесты корреляционной связи, а тем более об оценке ее существенности на основании только коэффициента Фехнера. При малом объеме исходной информации коэффициент Фехнера практически решает ту же задачу, которая ставится при построении групповых корреляционных таблиц, т.е. отвечает на вопрос о наличии и направлении корреляционной связи между признаками. В том случае, если построена корреляционная или же групповая таблица, дополнительный расчет коэффициента Фехнера не имеет практической ценности.

М. Кендэл предложил еще одну меру связи между переменными x и y — коэффициент корреляции рангов Кендэла(τ):

$$\tau = \frac{2S}{n(n-1)},$$

где $S = P + Q$.

Для вычисления τ надо упорядочить ряд рангов переменной x , приведя его к ряду натуральных чисел. Затем рассматривают последовательность рангов переменной y .

Для нахождения суммы S находят два слагаемых P и Q . При определении слагаемого P нужно установить, сколько чисел, находящихся справа от каждого из элементов последовательности рангов переменной y , имеют величину ранга, превышающую ранг рассматриваемого элемента.

Поскольку коэффициенты корреляции рангов могут изменяться пределах от -1 до $+1$ (как и линейный коэффициент корреляции), по результатам расчетов коэффициента Спирмэна можно предположить наличие достаточно тесной прямой зависимости между оценками экспертов на стадии предвыборной кампании и результатами выборов. Однако нельзя не учесть то обстоятельство, что ранговый коэффициент был рассчитан по небольшому объему исходной

информации ($n = 10$). Не является ли отличие рангового коэффициента от нуля лишь результатом случайных совпадений оценок экспертов с результатами выборов по данным малого числа отобранных депутатов, можно ли распространить полученные выводы на генеральную совокупность?

Для совокупностей небольшого объема ($n < 30$) распределение рангового коэффициента корреляции не является нормальным, поэтому нецелесообразно использовать значения t по нормированной функции Лапласа для проверки гипотезы о величине рангового коэффициента корреляции. В справочной литературе приводится таблица предельных значений коэффициентов корреляции рангов Спирмэна при условии верности нулевой гипотезы об отсутствии корреляционной связи при заданном уровне значимости и определенном объеме выборочных данных.

Могут встретиться случаи, когда невозможно установить ранговые различия нескольких смежных значений. В этих случаях принято брать средний ранг (даже если он будет дробным числом) и полученный средний ранг приписывать каждому из таких значений, т.е. переходить к матрице переформированных рангов. Например, двум факторам один из экспертов приписывает одинаковый ранг 3. Тогда каждому из факторов присваивается ранг 3,5, так как они поделили между собой третье и четвертое места $(3 + 4)/2$, а фактору, имевшему ранг 4, присваивается ранг 5 и т.д.

Если определяется теснота связи между k -м и l -м признаками, в рядах значений которых имеется соответственно q_i и g_l групп объединенных рангов, то формула коэффициента корреляции рангов Спирмэна примет вид:

$$\rho = \frac{\frac{n^3 - n}{6} - (T_k - T_l) - \sum_{i=1}^n d_i^2}{\sqrt{\left(\frac{n^3 - n}{6} - 2T_k\right) \times \left(\frac{n^3 - n}{6} - 2T_l\right)}}$$

, где $T_k = \sum_{i=1}^q \frac{t_{k_i}^3 - t_{k_i}}{12}$; $T_l = \sum_{i=1}^g \frac{t_{l_i}^3 - t_{l_i}}{12}$; t_{k_i} и t_{l_i} определяют количество единиц в i -й группе объединенных рангов соответствующего признака. Скорректированная формула для вычисления коэффициента Корреляции рангов Кендэла будет иметь вид:

$$\tau = \frac{S}{\sqrt{\left(\frac{n(n-1)}{2} - V_k\right) \times \left(\frac{n(n-1)}{2} - V_l\right)}}$$

где $V_k = \sum_{i=1}^q \frac{t_{k_i}(t_{k_i} - 1)}{2}$; $V_l = \sum_{i=1}^g \frac{t_{l_i}(t_{l_i} - 1)}{2}$.

Моделирование временных рядов

Эконометрическую модель можно построить, используя два типа исходных данных: 1) данные, характеризующие совокупность различных объектов в определенный момент (период) времени; 2) данные, характеризующие один объект за ряд последовательных моментов (периодов) времени.

Модели, построенные по данным первого типа, называются пространственными моделями. Модели, построенные по данным второго типа, называются моделями временных рядов.

Временной ряд - это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы: 1) факторы, формирующие тенденцию ряда; 2) факторы, формирующие циклические колебания ряда; 3) случайные факторы.

При различных сочетаниях этих факторов зависимость уровней ряда от времени может принимать разные формы. Во-первых, большинство временных рядов экономических показателей имеют тенденцию, характеризующую совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. По всей видимости, эти факторы, взятые в отдельности, могут оказывать разнонаправленное воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию. Во-вторых, изучаемый показатель может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей зависит от времени года. При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка, а также с фазой бизнес-цикла, с фазой, в которой находится экономика страны. Некоторые временные ряды не содержат тенденции и циклическую компоненту, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты.

Очевидно, что реальные данные не соответствуют полностью ни одной из описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных колебаний и случайной компоненты. В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется аддитивной моделью временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется мультипликативной моделью временного ряда. Основная задача эконометрического исследования отдельного временного ряда - выявление и придание количественного

выражения каждой из перечисленных выше компонент, с тем чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

При наличии тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих значений. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Количественно ее можно измерить с помощью парного линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называется лагом. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило «максимальный лаг должен быть не больше $n/4$ ».

Отметим два важных свойства коэффициента автокорреляции. Во-первых, он строится по аналогии с линейным коэффициентом корреляции и, таким образом, характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю. Во-вторых, по знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом они могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют автокорреляционной функцией временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется коррелограммой. Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, следовательно, лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т.е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка, ряд содержит циклические колебания с периодичностью в моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать

предположение относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний и имеет структуру, сходную со структурой ряда, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ. Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты и циклической (сезонной) компоненты. Аналогично, если, например, при анализе временного ряда наиболее высоким оказался коэффициент автокорреляции второго порядка, ряд содержит циклические колебания с циклом, равным двум периодам времени, т.е. имеет пилообразную структуру.

Одним из наиболее распространенных способов моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называют аналитическим выравниванием временного ряда.

4. ЗАДАНИЯ ДЛЯ КУРСОВОЙ РАБОТЫ

(Задачи корректируются в соответствии с двумя последними цифрами зачетной книжки студента)

Задание № 1

Для решения Задания №1 исходные данные представленные в Таблице № 1 (Приложение 1) необходимо изменить, прибавив к каждому показателю (для задачи №1) X_1 и X_2 две последние цифры зачетной книжки студента, например, 92 (95+92 и т.д.) показатель Y оставить без изменения. В случае если номер заканчивается на цифру 0, то для расчета берется предыдущие два числа, если номер заканчивается, например на 906 для расчета берется предыдущая цифра, соответственно число 96 и т.д. Две последние цифры зачетной книжки прибавляемые к каждому числу являются номером выполняемого варианта, который в обязательном порядке указывается на титульном листе к курсовой работе установленного образца (см. прил. 2).

По вновь образованным данным построить:

1. в рамках значений Y , X_1 и Y , X_2 отдельно парную корреляционную модель, описывающую влияние на уровень производительности труда:
 - а) текучести рабочих кадров X_1
 - б) уровня механизации строительно-монтажных работ X_2 .
- В рамках каждой зависимости:
 - 1.1. Оценить значимость полученного коэффициента корреляции, проверив справедливость нулевой гипотезы.
 - 1.2. Оценить значимости уравнения регрессии с помощью критерия Фишера.
2. Построить множественную корреляционную модель описывающую совокупное влияние на уровень производительности труда двух факторов: текучести рабочих кадров X_1 и уровня механизации строительно-монтажных работ X_2 .
3. Оценить степень тесноты связи посредством расчета коэффициента Фехнера в рамках значений Y , X_1 и Y , X_2 ;
4. Рассчитать коэффициент корреляции рангов (коэффициент Спирмена) для оценки тесноты связи между показателями X_1 и Y и показателями X_2 и Y .

5. Рассчитать коэффициент конкордации в рамках значений X_1, X_2, Y .
6. Рассчитать частные коэффициенты эластичности в рамках построенной множественной корреляционной модели, взяв коэффициенты регрессии по соответствующим факторным признакам вспомогательная таблицы в рамках решения п.2. Задачи 1.

По рассчитанным показателям сделать соответствующие выводы.

Пояснения к Заданию № 1

Введены следующие обозначения: Y - производительность труда на одного работающего, тыс. руб. / год; X_1 - текучесть рабочих кадров, % ; X_2 - уровень механизации строительно-монтажных работ, %.

Задание № 2

К каждому показателю Выпускники Вуза X , Выпускники Вуза Y прибавить число сформированное двумя последними цифрами зачетной книжки, решить задачу с использованием вновь образованных данных.

Исследуем связь между трудоустройством выпускников вузов г. Воронежа и уровнем их образования за 2018 год:

Уровень образования	Количество выпускников	Из них	
		Выпускники Вуза X	Выпускники Вуза Y
среднее профессиональное	-	373	489
Высшее	-	396	438
Итого	-	-	-

Рассчитать коэффициент ассоциации (K_a) и коэффициент контингенции (K_k). Сделать развернутый вывод.

Задание № 3

К каждому показателю прибавить число, сформированное двумя последними цифрами зачетной книжки, например, 20 ($46+20$, $89+20$ и т.д.) решить задачу с использованием вновь образованных данных.

Для изучения влияния условий производства на взаимоотношения в коллективе было проведено выборочное обследование рабочих, ответы которых распределены следующим образом:

Условия производства	Взаимоотношения в коллективе			
	хорошее	удовлетворительное	неудовлетворительное	итого
Соответствуют требованиям	46	70	89	-
не полностью соответствуют	45	60	19	-
не соответствуют	40	61	70	-
итого	-	-	-	-

Требуется охарактеризовать связь между исследуемыми показателями с помощью коэффициента взаимной сопряженности Пирсона и Чупрова, сформулировать вывод.

Задание № 4

К каждому показателю 2017, 2018, 2019 гг. прибавить число, сформированное двумя последними цифрами зачетной книжки, например, 20 (310,3+20 и т.д.) решить задачу с использованием вновь образованных данных

По станциям технического обслуживания легковых автомобилей города X имеются следующие данные:

Месяц	Число поступивших заявок		
	2017	2018	2019
01	410,3	313,6	514,0
02	411,1	314,3	514,7
03	411,5	314,4	515,1
04	412,0	314,6	515,6
05	412,6	315,6	516,0
06	416,0	317,1	517,4
07	415,9	316,9	528,2
08	416,2	317,0	518,4
09	416,4	361,5	517,8
10	415,2	361,0	517,5
11	415,0	314,9	517,0
12	412,8	313,8	516,5

Требуется на основе приведенных данных выявить наличие сезонной неравномерности, определить величину сезонной волны с использованием индекса сезонности. Сделать вывод.

Задание № 5

К каждому показателю прибавить число, сформированное двумя последними цифрами зачетной книжки, например, 20 (20+20, 43+20 и т.д.) решить задачу с использованием вновь образованных данных

С помощью бисериального коэффициента корреляции исследовать связь между возрастом и социальным положением основных категорий потенциальных эмигрантов.

Основные категории потенциальных эмигрантов	Возраст, лет				Всего, чел.
	До 30	30-40	40-50	50 и более	
Руководители, чел	20	45	109	16	-
Рабочие, чел	43	45	28	53	-
Итого:	-	-	-	-	-

Сделать вывод.

Задание № 6

К каждому показателю прибавить число, сформированное двумя последними цифрами зачетной книжки, например, 20 ($75,5+20$, $80,5+20$ и т.д.) решить задачу с использованием вновь образованных данных

Для организационно-строительной отрасли анализируют заработную плату Y сотрудников в зависимости от масштаба (количества сотрудников) организации X . Наблюдения по 30 случайно отобранным организациям представлены следующей таблицей:

Y						X
75,5	75,5	77,5	78,5	80,0	81,0	100
80,5	82,0	84,5	85,0	85,5	86,5	200
85,5	88,5	90,0	91,0	95,5	96,0	300
93	93,5	97,5	99,0	201,5	105,0	400
102,0	105,5	107,0	110,5	115,0	118,5	500

1. Постройте уравнение регрессии Y и X и оцените его качество;
2. Можно ли ожидать наличие гетероскедастичности в данном случае? Ответ поясните.
3. Проверьте наличие гетероскедастичности, используя тест Голдфелда – Кванта. Рекомендуется использовать разбиение, при котором $R=12$.
4. Если предположить, что гетероскедастичность имеет место и дисперсии отклонений пропорциональны значениям X , то какое преобразование вы предложите, чтобы получить несмещенные, эффективные и состоятельные оценки?
5. Постройте новое уравнение регрессии на основе преобразования осуществленного в предыдущем пункте, и оцените его качество.
6. Сравните результаты полученные в пунктах 1 и 5.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Баркалов С.А., П.Н. Курочка, Е.Ю. Шмелева Практикум по статистике. Воронеж, ВГАСУ, 2006
2. Баркалов С.А., Курочка П.Н., Курносков В.Б. Статистика. УМК. Воронеж: «Научная книга», 2010 – 728 с.
3. Баркалов С.А., Курочка П.Н., Шмелева Е.Ю. Практикум по статистике. Воронеж, ВГАСУ, 2010.
4. Баркалов С.А., Курочка П.Н., Перевалова О.С. Статистика. Практикум. Воронеж, ВГАСУ, 2016.
5. Практикум по теории статистики – под ред. проф. Р. А. Шмайловой. М.: Финансы и Эконометрика, 2001.
6. Гмурман В.И. Теория вероятностей и математическая Эконометрика, М.: Высшая школа, 2003.
7. Вентцель Е.С. Теория вероятностей– 4-е изд. – М.: Наука, 1969. – 576 с.
8. Цыпин, А. П. Эконометрика в табличном редакторе Microsoft Excel [Электронный ресурс] : лабораторный практикум / А. П. Цыпин, Л. Р. Фаизова. — Электрон. текстовые данные. — Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2016. — 289 с. — 978-5-600-01401-5. — Режим доступа: <http://www.iprbookshop.ru/71328.html>
9. Васильева, Э. К. Эконометрика [Электронный ресурс]: учебник для студентов вузов, обучающихся по специальностям экономики и управления (080100) / Э. К. Васильева, В. С. Лялин. — Электрон. текстовые данные. — М. : ЮНИТИ-ДАНА, 2017. — 398 с. — 978-5-238-01192-9. — Режим доступа: <http://www.iprbookshop.ru/71058.html>
10. Гусаров, В. М. Эконометрика [Электронный ресурс] : учебное пособие для студентов вузов, обучающихся по экономическим специальностям / В. М. Гусаров, Е. И. Кузнецова. — 2-е изд. — Электрон. текстовые данные. — М. : ЮНИТИ-ДАНА, 2017. — 479 с. — 978-5-238-01226-1. — Режим доступа: <http://www.iprbookshop.ru/71166.html>

ПРИЛОЖЕНИЕ 1

№ п/п	текучесть рабочих кадров, % , x1	уровень механизации строительно-монтажных работ, %, x2	производительность труда на одного работающего, тыс. руб. / год, у
1	0,15	95	10,3
2	0,18	97	9,6
3	0,19	94	8,6
4	0,44	61	4,7
5	0,35	60	6,3
6	0,28	65	5,4
7	0,23	57	6,5
8	0,36	58	5,1
9	0,42	60	6,2
10	0,26	65	5,3
11	0,23	53	5,8
12	0,37	65	5
13	0,57	50	5,1
14	0,37	55	4,3
15	0,28	55	4,6
16	0,24	75	6,3
17	0,28	80	7,7
18	0,27	81	7,8
19	0,29	83	7,9
20	0,30	84	8,1

ПРИЛОЖЕНИЕ 2

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное
образовательное учреждение высшего образования

**«Воронежский Государственный
Технический Университет»**

Факультет экономики, менеджмента и информационных технологий

Направление подготовки

Кафедра Управления

Курсовая работа

по дисциплине Экономико-статистические методы

Вариант

Выполнил: студент курса
группы

Проверил:

Работа защищена: «_»__2021 г.

С оценкой: _____

Подпись преподавателя: _____

Воронеж 2021

**ЭКОНОМИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ
(РАЗДЕЛ ЭКОНОМЕТРИКА)**

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к выполнению курсовых работ
для обучающихся всех направлений, специальностей,
форм обучения

Составитель
Свиридова Татьяна Анатольевна

Издается в авторской редакции

Подписано к изданию 18.11.2021.
Уч.–изд. л. 2,4 «С».

ФГБОУ ВО «Воронежский государственный технический университет»
394026 Воронеж, ул. 20-летия Октября, 84